Spatial autoregressive model for a Dirichlet distribution

T. Nguyen [a,b], S. Moka [a,d], K. Mengersen [c] and B. Liquet [a,b]

[a] *School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia*
[b] *Laboratoire de Mathématiques et de leurs Applications, Université de Pau et des Pays de l'Adour, Anglet, France*
[c] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*
[d] *School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia*
Email: teo.nguyen@students.mq.edu.au

Abstract: Compositional data are widely utilized in various fields, such as ecology, geology, economics, and public health, as they effectively represent proportions or percentages of different components in a whole. However, due to their relative nature and the constraint of lying on a simplex, traditional statistical methods are not directly applicable to compositional data (Aitchison 1982).

Spatial dependencies often exist in compositional data, particularly when the components represent different land uses or ecological variables. Spatial autocorrelation can arise from shared environmental conditions or geographical proximity. Therefore, it is essential to incorporate spatial information into the statistical analysis of compositional data to obtain accurate and reliable results.

To handle compositional data, the Dirichlet distribution is commonly used as its support is a compositional vector. Maier (2014) proposed a regression model for Dirichlet-distributed data, but this model does not consider spatial dependencies, which limits its applicability in spatial problems.

In this study, we introduce a spatial autoregressive model for Dirichlet-distributed data that incorporates spatial dependencies between observations. We develop a maximum likelihood estimator on a Dirichlet density function that includes a spatial lag term. To expedite computations, we compute the derivatives and Hessian matrix.

We compare this spatial autoregressive model with the same model without spatial lag and test both models on synthetic and real datasets. Different spatial weights matrices are employed to account for their effect on the synthetic dataset.

The results demonstrate that incorporating spatial dependencies can improve the performance of the model and confirm that the efficiency depends on the definition of the spatial weights matrix (Anselin 1988). By considering the spatial relationships among observations, our model provides more accurate and reliable results for the analysis of compositional data. Future research could further explore the application of the proposed model in different fields and investigate alternative spatial weights matrices for compositional data analysis in diverse spatial contexts.

REFERENCES

Aitchison, J., 1982. The statistical analysis of compositional data. Journal of the Royal Statistical Society: Series B (Methodological) 44.2, 139–160.
Anselin, L., 1988. Spatial econometrics: methods and models. Vol. 4. Springer Science & Business Media.
Maier, M., 2014. DirichletReg: Dirichlet regression for compositional data in R.

Keywords: *Compositional data, Dirichlet distribution, regression, spatial autoregressive model*